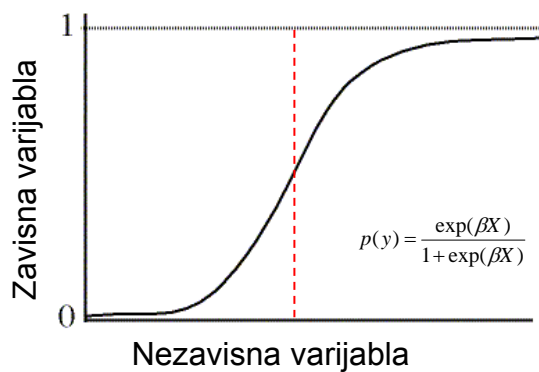


Tradicionalna metodologija scoringa u 17 slide-ova na primjeru churn-a



Kao odrediti “relevantnost atributa”

- Neki atributi “bolje” opisuju ciljnu varijablu od drugih
- npr. da li na *churn* više utječe dob, spol, regija, stručna sprema ili neka druga varijabla
- Zašto je važno odrediti stupanj relevantnosti atributa (obilježja) ?
- Uloga kombinatorike i kombinatorne eksplozije

Weights of evidence

$$WE = \log\left(\frac{\%non_churn_u_atributu}{\%churn_u_atributu}\right)$$

Tip kupca		Churn_flag		Weights of evidence
		N	Y	
		Column N %	Column N %	
Tip kupca	A	9.4%	19.9%	-0.754608573
	B	9.0%	16.5%	-0.60759719
	C	8.7%	14.8%	-0.53337319
	D	9.3%	14.7%	-0.453455997
	E	8.9%	13.5%	-0.421658213
	F	9.3%	7.3%	0.248786856
	G	9.4%	8.3%	0.12626254
	H	8.8%	5.0%	0.570260748
	I	9.1%	0.1%	4.156877462
	J	9.1%	0.0%	
	K	9.1%	0.0%	
		100.0%	100.0%	

© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com

Information value

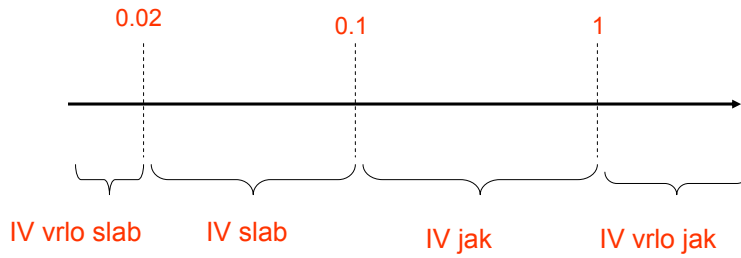
$$IV = \sum_{i=1}^n (\%non_churn_u_atributu_i - \%churn_u_atributu_i) * (we_i)$$

n= broj kategorija u atributu

Tip kupca		Churn_flag		Weights of evidence	Information value
		N	Y		
		Column N %	Column N %		
Tip kupca	A	9.4%	19.9%	-0.754608573	0.079617916
	B	9.0%	16.5%	-0.60759719	0.045651613
	C	8.7%	14.8%	-0.53337319	0.032617876
	D	9.3%	14.7%	-0.453455997	0.024221442
	E	8.9%	13.5%	-0.421658213	0.019603791
	F	9.3%	7.3%	0.248786856	0.005098149
	G	9.4%	8.3%	0.12626254	0.001401934
	H	8.8%	5.0%	0.570260748	0.021825239
	I	9.1%	0.1%	4.156877462	0.371764209
	J	9.1%	0.0%		
	K	9.1%	0.0%		
		100.0%	100.0%		IV= 0.601802168

© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com

Information value



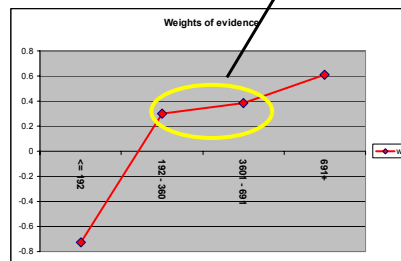
© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com

Optimalna kategorizacija

		oznaka_churna		we	IV
		N	Y		
		Column N %	Column N %		
Prosječni mjesečni iznos Izdanih faktura u €	<= 192	24.5%	50.6%	-0.7249	0.189299
	192 - 360	24.8%	18.4%	0.298033	0.019034
	360 - 691	25.3%	17.2%	0.385788	0.031214
	691+	25.4%	13.8%	0.612359	0.071256
		100.0%	100.0%		

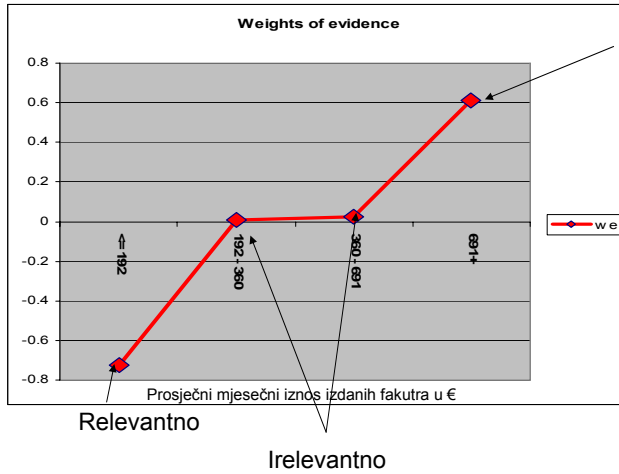
Grupiranje

Postojanje logike- we u negativnoj zoni implicira sklonost churnu, u pozitivnoj implicira sklonost "ne churnu"



© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com

“Dummy” varijable



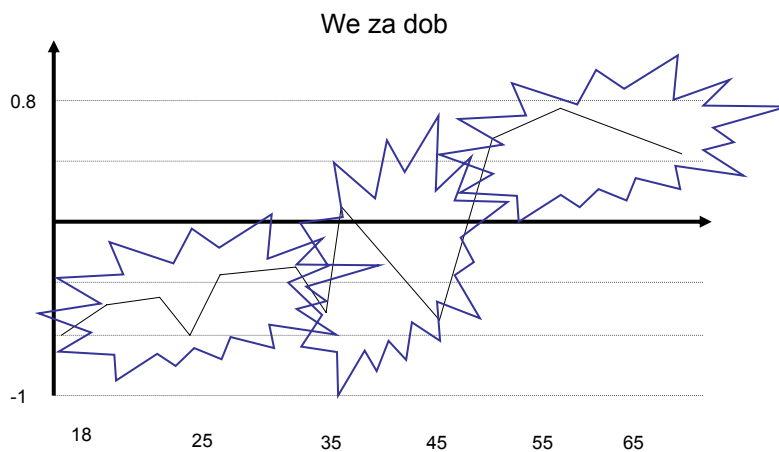
Relevantno

≤192	D1=1	D2=0
691+	D1=0	D2=1
192-360	D1=0	D2=0
360-691	D1=0	D2=0

Šifra kupca	D1	D2
1	0	1
2	1	0

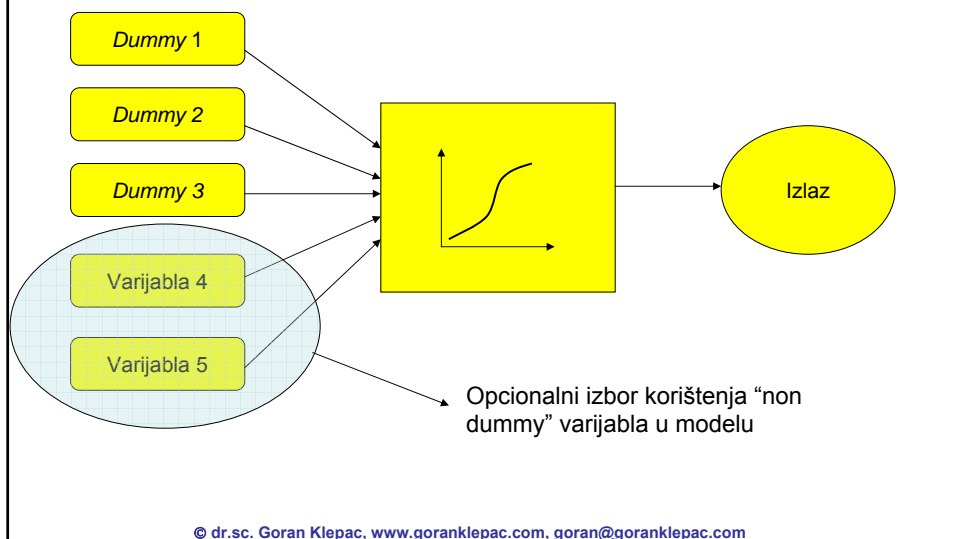
© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com

Situacija se komplicira kada ...



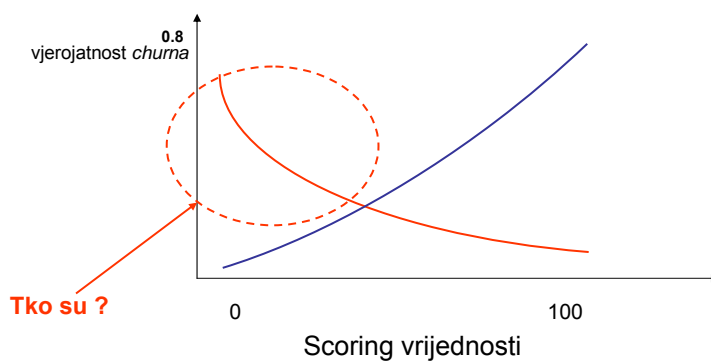
© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com

“Dummy” varijable i kreiranje modela



Klasični scoring modeli u službi predikcije *churna*

- Definicija klasičnog pristupa predikciji *churna*

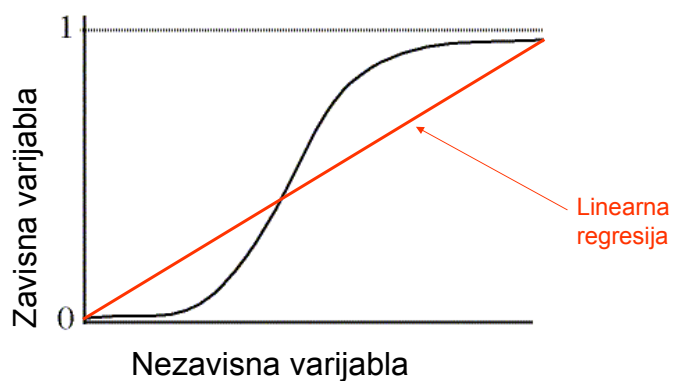


Izbor varijabli i veličina uzorka

- Kako i koje varijable izabrati ?
- Određivanje optimalne veličine uzorka
- Priprema varijabli i pretprocesiranje varijabli
- Selekcija metode
- Obrada
- Interpretacija rezultata
- Analiza pouzdanosti modela

© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com

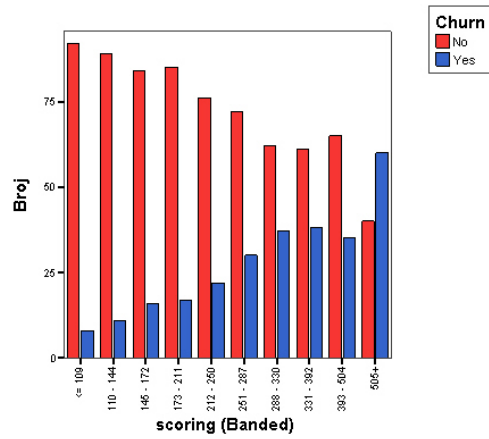
Logistička regresija



$$p(y) = \frac{\exp(\beta X)}{1 + \exp(\beta X)} \quad \text{Binarna logistička regresija}$$

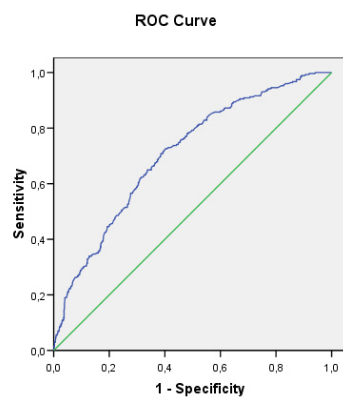
© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com

Scoring model



© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com

ROC krivulja



Diagonal segments are produced by ties.

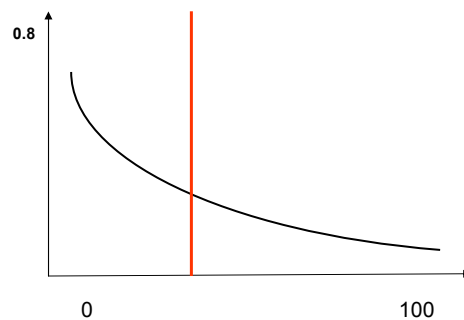
© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com

Korelatorni odnosi između nezavisnih varijabli

- Izbacivanje koreliranih varijabli
- Zadržavanje “relevantnijih” varijabli
- Sprečavanje *overfittinga*

© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com

Cut off strategija



© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com

Kalibracija

Stabilnost populacije:

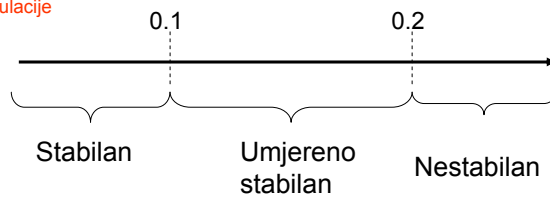
$$Index = \sum_{i=1}^n \ln\left(\frac{posto_{i(c)}}{posto_{i(d)}}\right) * (posto_{i(c)} - posto_{i(d)})$$

i – broj scorebandova

i(c)- tekući postotak unutar scorebanda

i(d)- novi postotak unutar scorebanda

Indeks stabilnosti
populacije



© dr.sc. Goran Klepac, www.goranklepac.com, goran@goranklepac.com